

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Virtual Screening for DNA Repair Inhibitors

Barakat, K.<sup>1</sup> and Tuszynski, J.<sup>1,2</sup>

<sup>1</sup>Department of Physics, University of Alberta,

<sup>2</sup>Department of Oncology, University of Alberta,  
Canada

## 1. Introduction

All living cells have a tendency to maintain their genomic stability with as few mutations as possible. This is of crucial importance to the normal function of cells in complex environments, correctly timed cell cycle progression, and a commitment to apoptosis when appropriate (Wood, et al., 2001). In this context, the balance between constancy and mutability in the context of genomic stability must be precisely regulated and controlled. To achieve this objective, a number of multiple and overlapping DNA repair pathways have been crafted within the cell (Harper & Elledge, 2007). Nevertheless, an elevated activity of these pathways could significantly decrease cancer cells' sensitivity to many known anticancer agents and, consequently, increase their antitumor drug resistance. This unforeseen role stems from the fact that most cancer chemotherapy in clinical use today, directly or indirectly damage DNA by causing single- or double-stranded DNA breaks or by interfering with the functions of crucial DNA interacting proteins. As a natural cellular response, following the detection of damage, DNA repair pathways attempt to restore the genome and restore the normal state of the cell. During this course, the cell's fate is mainly determined by the effectiveness of DNA repair mechanisms which allow the cell to survive or, if the damage is too heavy, induce apoptosis, causing the cell to die (Harper & Elledge, 2007). Consequently, to improve existing cancer therapies, DNA repair pathways have been considered as novel therapeutic targets. Several DNA repair inhibitors have been reported, some of which have been recently proven to be successful (Damia & D'Incalci, 2007).

This review paper focuses on our efforts directed at *in silico* searches for inhibitors of proteins that control the DNA repair circuitry. The targets chosen here play critical roles in tumor cell initiation and progression, hence their regulation offers promise for the improvement of current cancer therapy. Two of these targets are DNA repair proteins that are directly linked to the hallmark "relapse" or "drug resistance" phenomena. These are Excision Repair Cross-Complementation Group 1 (ERCC1) (Kang, et al., 2006), and DNA polymerase beta (pol  $\beta$ ) (Parsons, et al., 2004). The former is a key player in Nucleotide Excision Repair (NER), while the latter is the error-prone polymerase of Base Excision Repair (BER). The third target is p53 (Teodoro, et al., 2007), a so-called guardian protein of the genome that is inactivated in more than half of all human cancers investigated. An additional aim of this review is to share with the reader our experience as a computational drug discovery group by describing the virtual screening protocol we have developed in order to successfully address these biological problems. This chapter is divided into two main sections. The first gives a description of the computational workflow that we typically

follow in our virtual screening tasks. The second is a summary of our findings for the individual targets listed above.

## 2. An improved virtual screening (VS) protocol

Fig. 1 illustrates the essential steps taken in order to execute the virtual screening (VS) protocol used in our lab. In a nutshell, the developed protocol employs molecular docking, molecular dynamics simulations and clustering techniques to filter a given library of compounds for inhibitors of a particular target. The concepts behind VS and other computational tools are described elsewhere (Limin, et al., 2011); (Stahura & Bajorath, 2004). However, a detailed description and rationale behind each step of this workflow are summarized below. Except for a few steps that need careful preparation, the whole process has been automated. It starts with a collection of 3D structures of ligands and a well-prepared target structure. It finally yields a set of top hit structures in their preferred binding modes with the target. Although the following steps were applied to the three specific targets described in the following section, the procedure is general and the same method is applicable to almost any bio-molecular target.

### 2.1 Target preparation

#### 2.1.1 Primary assessment of target structure

In general, the downloaded “crude” crystal structure of a target contains many details that must be taken into account. This includes non-standard amino acids; co-factors; other small molecules that are present due to the crystallization process; ions and co-crystallized water molecules. For most small molecules like polyethylene glycol, it is advisable to remove them from the structure, since they are not included in the native form of the target but were required for the crystallization process. Moreover, non-standard amino acids must be carefully assessed and modeled. In many protein structures, these unusual amino acids lack several atoms because most structure handling packages do not check automatically inspect for them. Their parameters must be appended to the used Force Field (FF) before starting further simulations. Co-factors, ions and co-crystallized water molecules should be included within the simulated structure.

Water molecules that are located close to or within the binding site can mediate several interactions with the ligands. However, it is important to find out which water molecules are conserved within these regions. Any unpreserved (misplaced) water molecule can obstruct the docking simulation and lead to incorrect results. One way to identify important water molecules is to compare several crystal structures of the same target (if applicable) and choose the water molecules to be kept during the docking procedure. When a limited number of target structures is available, it is important to run different docking exercises by removing/keeping these water molecules and selecting the cases that lead to realistic and favorable binding modes.

An additional decision-making tool for the selection of water molecules is to use prediction software packages (e. g. ConSolv 1.0) that check whether a bound water molecule is likely to be conserved or displaced in other, independently solved crystallographic structures of the same target.

Finally, it is necessary to verify that no parts of the protein structure are missing. These missing residues are usually mentioned at the header of the Protein Data Bank (PDB)-file and must be added and relaxed within the target structure. Regarding the targets that were studied in this work, all missing amino acids were distant away from the binding site.

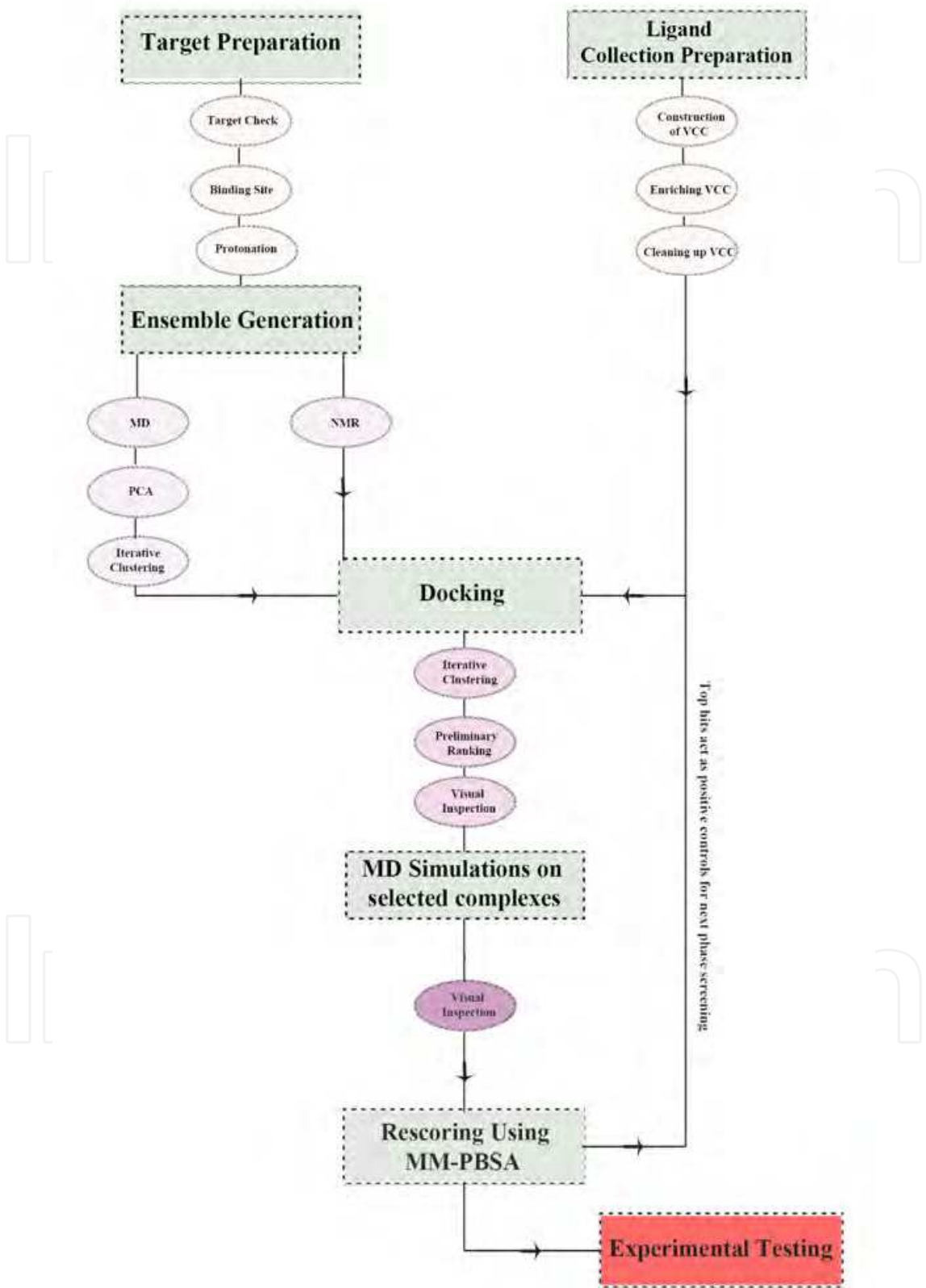


Fig. 1. Description of the implemented computational workflow. See text for more details.

Nonetheless, we added and relaxed them using molecular dynamics (MD) simulations before running the docking experiments (see below).

### 2.1.2 Identifying the binding site

The starting point of any VS study is the identification of the binding site within the target protein. This portion of the protein is directly related to the biological activity that needs to be regulated. At this stage, it is important to consult previously published work and determine if there are any known active compounds that bind to the target protein (positive controls) and to ascertain their binding location. If the binding site is not exactly known, however, there is a set of active molecular structures that exist, one should run a series of blind docking experiments until a suitable and experimentally verified binding site is found (Bennett, et al., 2010); (Hazan, et al., 2008). Regarding the three targets that we focus on here, the binding sites were accurately known, mainly because they are protein-protein interaction sites (e.g. ERCC1-XPA and p53-MDM2/MDM4), or protein-DNA binding sites (e.g. DNA-pol  $\beta$ ) where crystal structures of the interacting subunits are available.

### 2.1.3 Protonation states of charged residues

Proper adjustment of the protonation states, “the assignment of Hydrogen atoms” of the ionizable groups, contained by the target structure is important for any successful VS simulation. These residues play key roles in inter-protein, protein-solvent and protein-ligand interactions. The protonation states can be determined by predicting the pKa value of charged residues and comparing it to the pH value at which the simulation is performed. In this work, all protonation states of ionizable residues were calculated using the software PROBKA and adjusted at physiological pH of 7.0 (H. Li, et al., 2005). PROBKA is a very fast and accurate method that relates the structure and environment of the charged residues to the change of the pKa values from their intrinsic ones. Once the protonation states have been decided, all hydrogen atoms are then added to the system according to a given force field (FF). For the three targets, the AMBER99SB FF was used (Hornak, et al., 2006). At this stage, the protein structure is ready for the docking or MD simulations.

## 2.2 Ligand collection preparation

In parallel with target preparation, the organization and cleaning up of the set of compounds is undertaken for *in silico* screening. Currently, there are many suitable, easy to access compound databases that contain millions of molecules spanning various structural families. Prior to any screening exercises, one should decide on a set of compounds to be filtered and build up a virtual compound collection (VCC) of compounds. This collection will be repeatedly used against many targets. A typical VCC should include marketed drugs, lead-like compounds, fragment structures, commercially available chemicals and other high-activity molecules. It is also important to represent these molecules in different protonation, stereo and conformational states. An effective VCC should be constructed from molecules that are suitable for further lead optimizations, after they show biological activity.

### 2.2.1 Construction of the VCC

Five different databases comprise the core of our VCC. These are the National Cancer Institute diversity set (NCIDS), the DrugBank database (Wishart, et al., 2006), subsets of the ZINC database (Irwin & Shoichet, 2005) and finally, the French national chemical library “la



Chimiothèque Nationale" (CN). Some of them are used in the first iteration of VS and others are retained for higher-order screening exercises.

The NCIDS is a collection of approximately 2,000 compounds that are structurally representative of a wide range of molecules, representing almost 140,000 compounds that are available for testing at the NCI. A number of its ligands contain rare earth elements and cannot be properly parameterized for docking experiments, leaving us with 1,883 compounds that can be actually used. We use a cleaned 3D version of the NCIDS formatted for use in AutoDock (the main docking program used by us) (Goodsell & Olson, 1990) and was prepared by the AutoDock Scripps team. What makes the NCIDS so valuable and extensively screened by many groups (even in HTS) is that its individual molecules have distinctive structures and are the cluster representatives of their parent families. Having screened and ranked the molecules, one can re-screen the subset of the representative structures, instead of screening the entire NCI set of compounds.

The DrugBank database is not only a set of molecules representing FDA-approved (and investigational) drugs, but also it is a unique bioinformatics and cheminformatics resource since it relates each drug to its target(s). It includes details about the different pathways, structural information and chemical characteristics of these targets and the way they are involved in a particular disease. This information is stored on a freely accessible website that is linked to other databases (KEGG, PubChem, ChEBI, PDB, Swiss-Prot and GenBank) and to a range of structure displaying applets. The DrugBank collection includes ~4,800 drug structures including >1,350 FDA-approved small molecule drugs, 123 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,243 experimental drugs. Once a hit is identified from this library, it is simply a drug. This means it overcomes many barriers of preclinical and clinical testing and development and can be directly tested for its novel biological activity. Moreover, a hit from this collection may explain a mysterious side effect that would not be discovered before its identification as a regulator of the examined target.

ZINC is a free database dedicated to VS. It includes more than 13 million purchasable compounds most of which are "drug-like" or "lead-like". These compounds are available in several 3D formats and compatible with several docking programs. The ZINC database has many other interesting features. For example, one can easily create a subset of the whole database with any given set of properties such as specific functional groups, molecular weight, and a calculated logP. Most of the compounds also exist in multiple protonation states suitable for different pH values, several tautomeric forms, all possible stereochemistries, and different 3D conformations. The database is also organized so that the origin of each molecule is known. That is, one can determine the vendor and original catalog number for each commercial source of a compound. Similarly to the DrugBank database, a molecule can be annotated for its function or activity. It also has a powerful web server that helps in searching, browsing, creating subsets, and downloading some or all of the molecules in the database.

The CN chemical library (~100,000 compounds) is a repository of all synthetic, natural compounds and natural extracts in the existing French public laboratories. This database is divided into two main categories. The first part includes information about all synthetic products, while the second contains the natural compounds and extracts. In this work, we used the whole CN database in our screening. In contrast to the previously mentioned databases, compounds in this library are represented by 2D SDF structures with no hydrogen atoms attached. This required a number of cleaning and preparation steps before using them in our VS simulations (see below).

### 2.2.2 Enriching the VCC core

This is where ligand-based methods come to play a significant role in the pre-screening process. Any molecule that is known to bind to the target-binding site can serve as a positive control. Such molecules can be identified through published articles or previous patents. Besides their function in directing and verifying the simulation parameters, they can be used as seeds in the searches for similar chemical structures to enrich the VCC. This is a crucial step, which should be taken even if the identified similar structures have been previously removed from the VCC in its early construction steps.

Following this strategy, we have used known inhibitors for the p53-MDM2 interaction (see section 3.2) and DNA pol  $\beta$  (see section 3.3) to enrich their representative VCCs. For the ERCC1-XPA interaction (see section 3.1), initially, there was no active compound confirmed to bind to the ERCC1 pocket. Hence, for the first round of VS, we started from scratch and did not apply this enrichment method. However, it was used in the second round of screening, after the first iteration identified a list of novel binders to the ERCC1 target.

### 2.2.3 Cleaning up the VCC

Having decided which collection of compounds to use in the screening process, one should spend time and effort to ensure the quality of the used ligand structures. As mentioned before, it is important to adopt proper protonation and conformational states for the ligands. For example, the original CN library of compounds is a collection of 2D structures with no hydrogen atoms. Ligands in this state are not suitable for docking using many of the popular docking programs. These software packages require 3D structures with proper placement of hydrogen atoms. One solution to this problem, which was followed in the ERCC1-XPA case (see section 3.1), is to use conversion software that can translate the 2D information into its 3D representative structure. Many of such programs are available (e.g. Open Babel and LigPrep from Schrödinger). We prefer LigPrep for this task because it produces structures with few errors compared to Open Babel, especially in bond connection and hydrogen atoms assignment.

## 2.3 Generation of an ensemble of target structures

Proteins are inherently dynamical macromolecules. Their dynamical behavior is essential in order to recognize and bind to other molecules inside the cell. Although many attempts have been made to partly include the flexibility of the molecular target within docking algorithms (Schneider & Bohm, 2002), there are still many barriers and challenges that impede progress in this field. One major challenge is the enormous number of conformations that are accessible to the target under equilibrium conditions. The range of these conformations is very wide and includes many local and global movements within the structure of the protein. These dynamical transitions can be as small as minor rotations of the side-chains or as large as the complete dislocation of domains within the same target. There are many crystal structures in the PDB that give evidence to this bizarre dynamical behavior. These conformational changes can be illustrated by comparing different crystal structures of the same target, especially, between its bound and unbound forms.

### 2.3.1 Hybrid MD-docking methods

One way to accommodate receptor flexibility and to offer more accurate scoring techniques is to implement a hybrid method between docking and MD simulations. Originally, the use

of MD simulations in VS studies was intended to create a set of receptor conformations (Broughton, 2000; Carlson, et al., 2000). However, it was always debatable whether to use structures derived from MD simulations or NMR data. In our opinion, if a reasonable ensemble of NMR structures exist, one should consider using them all, instead of running long MD simulations. However, if the VS exercise departs from a single X-ray crystal structure, it is important to generate such an ensemble using MD simulations.

In this context, a successful approach, reported by McCammon and his team, is the relaxed complex scheme (RCS)(Lin, et al., 2002; Amaro, et al., 2008). This method, illustrated in Fig. 2, forms the foundation of the VS protocol presented here. In the RCS approach, all-atom MD simulations (e.g., 2-5 ns simulation) are applied to explore the conformational space of the target, while docking is subsequently used for the fast screening of drug libraries against an ensemble of receptor conformations. This ensemble is extracted at predetermined time intervals (e.g., 10 ps) from the simulation, resulting in hundreds of thousands of protein conformations. Each conformation is then used as a target for an independent docking experiment.

### 2.3.2 Principle component analysis and sampling convergence

A typical MD trajectory displays the time dependence of atomistic Cartesian coordinates. Although the duration of the whole trajectory is typically very short (at best, on the order of hundreds of ns) compared to real life biological dynamics, it involves a huge number of snapshots that contain a mixture of fast and slow modes of motion. It is impossible to segregate or understand this mixed dynamics through simple analysis (e.g. visual inspection). However, covariance, or principle component, analysis (PCA) can break up these two types of motions and extract the essential dynamics (ED) spanned by the protein structure. This essential dynamics represents the collective movements that are directly linked to the function of the protein and are essential for its role. In fact, PCA transforms the original space of correlated variables from a large MD simulation into a reduced space of independent variables (Garcia, 1992); (Amadei, et al., 1993). For a typical protein, the system's dimensionality is thereby reduced from tens of thousands to fewer than fifty degrees of freedom.

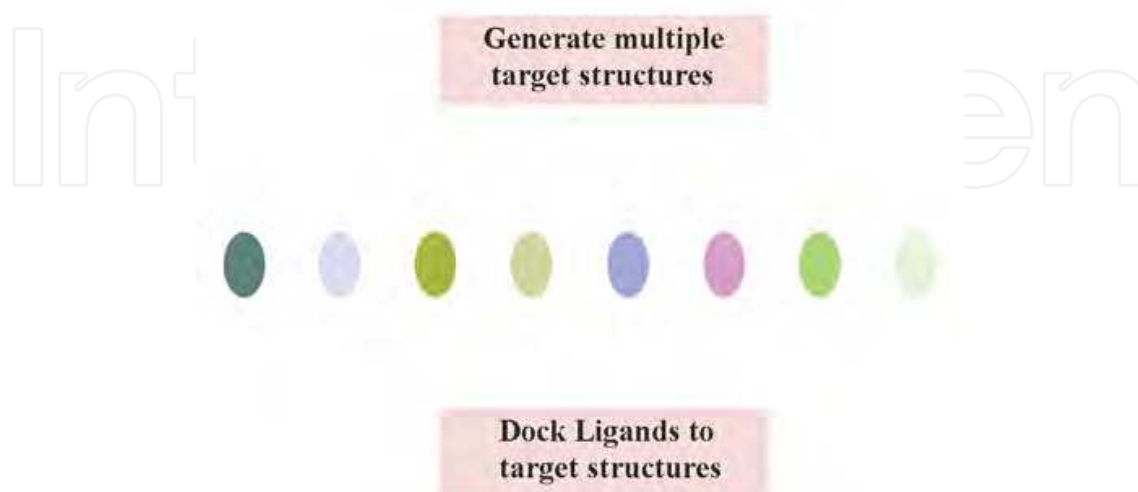


Fig. 2. Schematic illustration of the basic idea behind the relaxed complex scheme developed by McCammon et al.(Lin, et al., 2002).



To perform PCA for a subset of  $N$  atoms, the entire MD trajectory is RMSD fitted to a reference structure, in order to remove all rotations and translations. The covariance matrix can then be calculated from their Cartesian atomic co-ordinates as:

$$\sigma_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle \quad (1)$$

where  $r_i$  represents the three Cartesian co-ordinates ( $x_i, y_i$  or  $z_i$ ) and the eigenvectors of the covariance matrix constitute the essential vectors of the motion. It is generally accepted that the larger an eigenvalue, the more important its corresponding eigenvector in the collective motion. PCA can also be employed to predict the completeness of sampling during the MD simulation. This step is critical and was used in three cases to answer a very important question: when to stop the simulation and start extracting the dominant conformations of the protein? In this, we follow a method proposed by Hess (Hess, 2002) that divides an MD trajectory into separate parts, and their normalized overlap is calculated using the covariant matrices for each pair of parts:

$$\text{normalized overlap } (C_1, C_2) = 1 - \frac{\sqrt{\text{tr} \left( \left( \sqrt{C_1} - \sqrt{C_2} \right)^2 \right)}}{\sqrt{\text{tr}(C_1) + \text{tr}(C_2)}} \quad (2)$$

where  $C_1$  and  $C_2$  are the covariant matrices, and the symbol  $tr$  denotes the trace operation. If the overlap is 0, then the two sets are considered to be orthogonal, whereas an overlap of 1 indicates that the matrices are identical. In this context, for the three targets studied described below, the individual whole trajectories were divided into three parts and the normalized overlap between each pair was calculated to determine the completeness of sampling.

### 2.3.3 Iterative clustering to extract dominant conformations

Once a sufficient sampling is confirmed through the aforementioned PC calculations, clustering analysis is used to extract a set of target structures that represent dominant conformations. Unfortunately, there is no universally accepted clustering algorithm or parameters that can be used to extract all information contained within the MD simulation. However, recent studies suggest that a number of clustering algorithms, such as average-linkage, means and self-organizing maps (SOM) can be used accurately to cluster MD data (Shao, et al., 2007). In this work, the clustering quality was anticipated by calculating a number of clustering metrics. These metrics can reveal the optimal number of clusters to be extracted and their population size. These are the Davies-Bouldin index (DBI) (Davies DL & Bouldin DW, 1979) and the "elbow criterion" (Shao, et al., 2007). A high-quality clustering scheme is correlated with high DBI values. On the other hand, using the elbow criterion, the percentage of variance explained by the data is expected to plateau for cluster counts exceeding the optimal number of clusters. Using these metrics, by varying the number of clusters, one should expect for adequate clustering, a local minimum for DBI and a horizontal line for the percentage of variance explained by the data. Fig. 3 describes an example of such calculations.

Our implementation employs an iterative clustering algorithm using the above-mentioned hypothesis. The procedure is established as an in-house code using the PTRAJ utility of

AMBER10 (Case, et al., 2005). A modified version of the code is also used to cluster the docking results. MD trajectories' clustering runs the average-linkage algorithm for a number of clusters ranging from 5 to 150 clusters. Structures are extracted at 2 ps intervals over the entire simulation time. In order to remove the overall rotation and translation, all C<sub>α</sub> atoms are fitted to the minimized initial structure. RMSD-clustering is performed on the residues contained in the investigated binding sites. These residues are clustered into groups of similar conformations using the atom-positional RMSD of the entire amino acid, including side chains and hydrogen atoms, as the similarity criterion. The centroid of each cluster, the structure having the smallest RMSD to all members of the cluster, is chosen as the cluster representative structure and the most dominant structures are used as rigid templates for the ensemble-based docking experiments (see below).

## 2.4 Docking ligands to the ensemble of target structures

As stated above, the outcome of the iterative clustering step is an ensemble of protein structures that are used as targets for docking. The main docking program that was used for the three cases analyzed was AutoDock version 4 (Garrett MM, et al., 1999). AutoDock is one of the most popular docking packages that utilize different conformational search methods, including Simulated Annealing (SA), traditional Genetic Algorithm (GA), and Lamarckian Genetic Algorithm (LGA). Here, we use the LGA approach. The approach is well-described in the original paper by Morris et al. (Garrett MM, et al., 1999).

### 2.4.1 Automated clustering of docked poses

AutoDock can cluster output poses into subgroups depending on their RMSD values referred to a reference structure. Although this approach is widely used, the number of clusters and the population size of each cluster strongly depends on the RMSD cut-off used. Consequently, it is impossible to predict the optimal cut-off for the RMSD in order to produce a clustering pattern with the highest confidence. This motivated us to use an alternative approach when clustering the docked ligand structures. In fact, we extended and automated the clustering methodology that was used in section 2.3.3 to couple the elbow criterion (Shao J, et al., 2007) with the clustering module of PTRAJ (Case, et al., 2005). This method exploits the fact that the percentage of variance exhibited by the data ( $\lambda$ ), is expected to plateau for cluster counts exceeding the optimal number.

The percentage of variance is defined by:

$$\lambda = \frac{SSR}{SST} \quad (3)$$

where (SSR) is the sum-of-squares regression from each cluster summed over all clusters and (SST) is the total sum of squares. Here, we used the SOM algorithm to cluster the docking results. This modified clustering program increases the number of clusters required until the percentage of variance exhibited by the data ( $\lambda$ ) plateaus. The convergence of clustering can be determined by calculating the first and second derivatives of the percentage of variance with respect to the clusters number ( $d\lambda/dN$  and  $d^2\lambda/dN^2$ ) after each attempt to increase the cluster counts. The clustering process then stops at an acceptable value for these derivatives that is close to zero. In this way, the clustering procedure depends only on the system itself and adjusts itself to arrive at the optimal clustering pattern for that specific system.

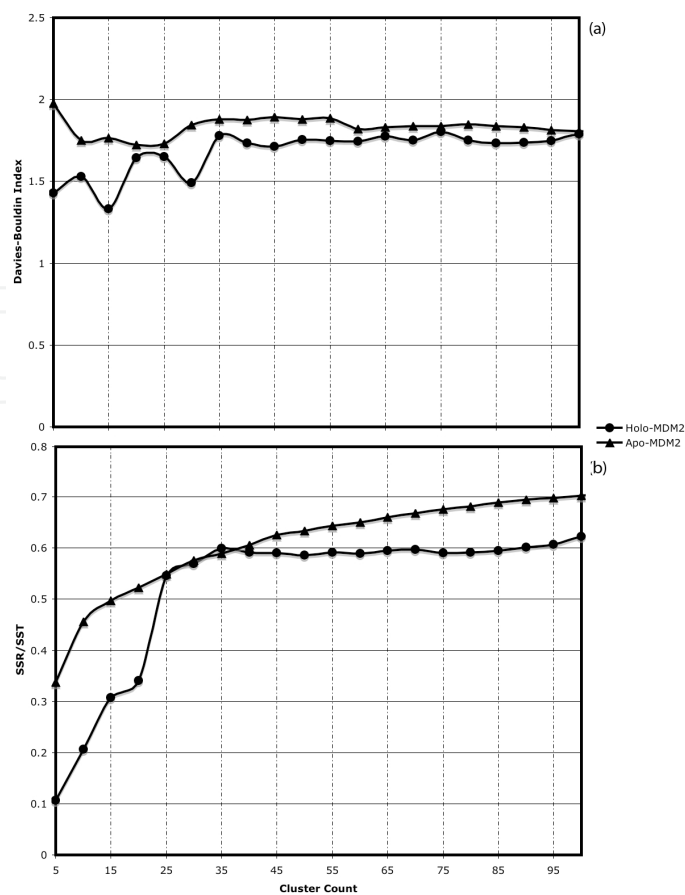


Fig. 3. Clustering Analysis. A high-quality clustering is obtained when a local minimum in DBI correlates with saturation in the SSR/SST ratio (K. Barakat, et al., 2010).

2.4.2 Preliminary ranking of docking results

The VS protocol then sorts the docking results by the lowest binding energy of the most populated cluster. The compounds can also be ranked using their weighted average binding energies according to the following formula:

$$\text{Weighted Average Binding Energy (WABE)} = \sum_i^M \text{percent ditribution}(i) \times \text{binding energy} \tag{4}$$

where *i* is the index number of each ensemble cluster, whose percent distribution sums up to 100% and *M* is number of different structures included in the ensemble. The VS protocol only considers a compound among the top hits if the most populated cluster from any of the VS experiments includes at least 25% of all docked conformations. The top N hits of the combined docking runs construct an irredundant set of promising compounds that are used for further analysis. In this work, a typical preliminary set (N) includes from 200 to 500 compounds.

2.4.3 Visual inspection and selection of a focused set of hits

Visual inspection of the preliminary set of hits is necessary before proceeding to the later computationally rigorous steps. Although this step involves more human intervention, it

assures the quality of the docking results, which are not precise in terms of ranking or the final selection of binding geometries.

## 2.5 Molecular dynamics simulations on selected hit-target complexes

There are many factors (e.g. water content, protein flexibility, etc.) that are not well characterized within the docking context (Warren, et al., 2006). During this step, the VS protocol aims at accounting for these factors by performing MD simulations. Each simulation starts from the final docked structure. The important aspect at this stage is the solvation of the docked models. It is generally accepted that water molecules are not only involved in solvation/desolvation of the protein-ligand complexes, but also mediate their interactions and help in generating more suitable binding modes. MD simulations relax the structures, rearrange water and ion molecules and generate trajectories that are used during the next step of binding energy calculations. The output obtained from this step is a set of snapshots representing the trajectory of the MD simulations for each complex. Although this procedure requires extensive computational resources, it tends to improve the protein-ligand interactions and enhance their molecular complementarity.

## 2.6 Rescoring of hits using the MM-PBSA method

Besides using MD simulations to refine the docked structures, another essential constraint for a successful VS experiment is to accurately predict the binding energies. To correctly perform this task, we need to move away from simple docking scoring methods. However, we are also restricted by the need for a fairly fast method that can be applied to many systems at a reasonable computational cost. In this context, the VS protocol utilizes a fast and efficient scoring method to suggest the final ranked set of top hits. This method is the molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) technique. The method was initially proposed by Kollman et al. (Kollman, et al., 2000) and it combines molecular mechanics with continuum solvation models. The method has been extensively tested on many systems and shown to reproduce, with an acceptable range of accuracy, experimental binding data. It was also validated as a VS refining tool and revealed excellent results in predicting the actual binding affinities and in discriminating true binders from inactive (decoy) compounds (Abagyan & Totrov, 2001; Schneider & Bohm, 2002; Shoichet, et al., 1993). Its main advantages are the lack of adjustable parameters and the option of using a single MD simulation for the complete system to determine all energy values.

In this work we used the MM-PBSA method as implemented in AMBER. The total free energy is the sum of average molecular mechanical gas-phase energies ( $EMM$ ), solvation free energies ( $G_{solv}$ ), and entropy contributions ( $-TS_{solute}$ ) of the binding reaction:

$$G = EMM + G_{solv} - TS_{solute} \quad (5)$$

The total molecular mechanical energies can be further decomposed into contributions from electrostatic ( $E_{ele}$ ), van der Waals ( $E_{vdw}$ ) and internal energies ( $E_{int}$ ):

$$\Delta G^o = \Delta G_{gas}^{protein-ligand} + \Delta G_{solv}^{protein-ligand} - \left\{ \Delta G_{solv}^{ligand} + \Delta G_{solv}^{protein} \right\} \quad (6)$$

Furthermore, the solvation free energy can be expressed as a sum of non-electrostatic and electrostatic contributions:

$$\Delta G_{solv} \approx \Delta G_{solv}^{nonele} + \Delta G_{solv}^{ele} \quad (7)$$

The non-electrostatic part was approximated by a linear function of the (SASA). That is:

$$\Delta G_{solv}^{nonele} = \gamma \times SASA, \text{ where } \gamma = 7.2 \text{ cal/mol/\AA}^2 \quad (8)$$

### 3. Targeting DNA repair proteins

Below is a description of our target proteins and a summary of the results we found after applying the above-mentioned VS protocol to regulate their activity.

#### 3.1 Case I: inhibitors of the ERCC1-XPA interaction

Platinum-based cancer therapy is one of the most efficacious treatments for many cancer types including testes, ovary, head, neck, and lung cancers (Boulikas & Vougiouka, 2003). While generally efficient at inducing apoptosis, acquired resistance to platinum compounds has limited their efficacy and therefore reduced successful clinical use of these agents (McGuire & Ozols, 1998). One way of acquiring this type of drug resistance is the remarkable high activity of the Nucleotide Excision Repair (NER) pathway (Rabik & Dolan, 2007).

Although many proteins are involved in the NER machinery, only the over-expression ERCC1 (a 33-kDa protein) correlates with the augmented platinum resistance. This spectacular conclusion has been reached from several independent clinical trial investigations on ovarian (Kang, et al., 2006), colorectal (Shirota, et al., 2001), and non-small cell lung cancer (Lord, et al., 2002). ERCC1 forms a tight heterodimer endonuclease complex with XPF. At the final stages of NER, the ERCC1-XPF enzyme cleaves the damaged DNA strand at the phosphodiester bonds on the 5' side of the damage. Prior to the incision step in NER, the ERCC1-XPF endonuclease is recruited to the damaged DNA site through a secondary interaction between ERCC1 and XPA (L. Li, et al., 1994). This protein-protein interaction is necessary for a functional NER mechanism.

The NMR crystal structure was resolved by Tsodikov's group (Tsodikov, et al., 2007). The critical residue-residue interactions as determined through our binding energy predictions are shown in Fig. 4. A 14-residue peptide from XPA that includes three essential consecutive glycines (residues 72–74) is buried within a hydrophobic cleft within the central domain of ERCC1. This peptide has two critical characteristics (Tsodikov, et al., 2007). First, it is necessary and sufficient for binding to ERCC1. Second, and more importantly, it can compete with the full-length XPA protein in binding to ERCC1 and disrupting NER in vitro. Moreover, there is no other cellular function beyond NER that has been observed for XPA (Rosenberg, et al., 2001). These observations, coupled with the available crystal structure of this interaction make ERCC1 and XPA an extremely attractive target for computationally based development of small molecule inhibitors that are targeted for use in combination therapies involving cisplatin.

##### 3.1.1 Two-stage filtering procedure

Following the VS protocol described in section 2, we carried out a two-stage virtual screening procedure. Top hits from the first iteration were used as positive controls for the second. First, we screened relatively small chemical libraries from the constructed VCC (see section 2.2.1). The objective at this stage was to discover novel lead compounds that can be used as positive controls for subsequent screening simulations and experiments.



Furthermore, we wished to construct a pharmacophore model that can guide future research toward discovering potent and specific inhibitors of this interaction. A detailed description of the parameters used and the results obtained in this study can be found in the published manuscript (K. H. Barakat, et al., 2009). Here, we summarize the major steps and report important findings.

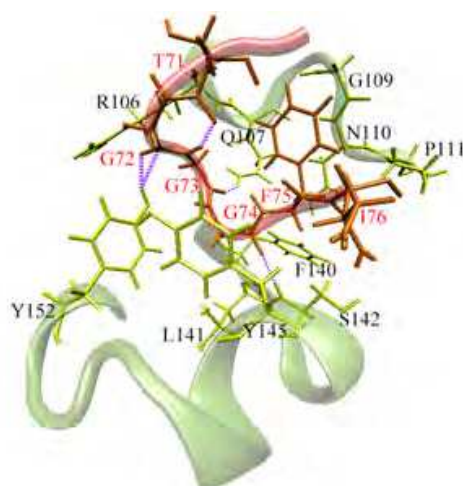


Fig. 4. An illustration of the XPA-ERCC1 protein-protein interaction. The binding between ERCC1 (teal) and XPA (red).

Prior to the screening process, it was important to determine precisely the key residues that mediated the ERCC1-XPA interaction (see Fig. 4). To do so, we used binding energy decomposition analysis using the MM-PBSA method in order to mark the essential residues from both sides. Residues within ERCC1 highlighted the binding site of the target. Following that, we extended the MD simulations up to 50 ns and extracted 6 dominant structures using clustering analysis. The length of the simulation was determined by convergence of the normalized overlap using PCA (see section 2.3.2 for details). Beside the 6 extracted structures we added 2 more protein conformations, one for a minimized initial structure and the other was an equilibrated conformation obtained from PCA. We also tried two docking alternatives. The intention was to compare docking against one target structure with flexible side chains to docking against eight rigid protein conformations. Flexible docking exploited the findings of the earlier analysis of binding energy decomposition, where we found Y145 to contribute more than 25% of the overall binding energy (data not shown). Hence, we decided to allow the full side-chain flexibility for this residue in docking to the minimized initial structure. In the other approach, we docked the two compound libraries to the eight rigid models of the target. We also used 3 different methods to rank the compounds; all of them were based on the AutoDock scoring function. In all ranking methods we used only the binding modes that possessed the largest docking cluster. The first ranking method was based on the minimum energy of the flexible docking run. The second used the average binding energy from the eight different rigid simulations. The final ranking was based on the weighted average binding energy using Equ. 4.

The binding mode for three selected top hits within their most favored binding site conformations is shown in Fig. 5. Electrostatic surface maps are included to provide an additional perspective of the charge distribution in the ERCC1 cavity. The binding cleft is mainly positively charged with small negatively charged spots on boundaries of the binding

site. This electrostatic potential distribution indicates that the binding site may exhibit a weak positive electrostatic potential. Although, the charge distribution changed slightly between the two representative binding sites indicating the perseverance of its overall shape, the positive potential is apparent on closed conformation. It is worth mentioning that thirty compounds out of this study were experimentally tested and two of them (including the one shown in Fig. 5-A) exhibited positive activity. These two identified top hits were then used as positive controls during the second stage of VS (see below).

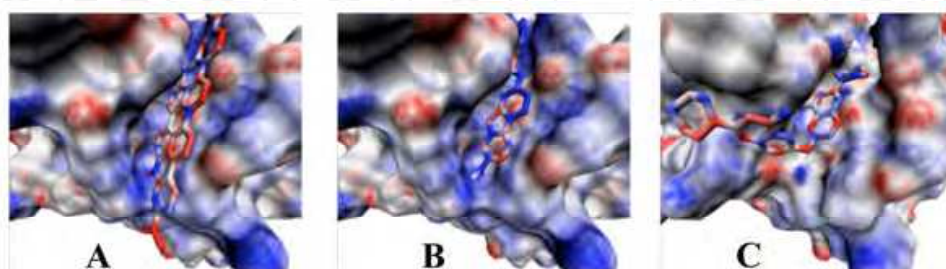


Fig. 5. Three selected hits within their preferred binding site conformations.

Two important lessons were learnt from the first stage and guided us toward more successful outcomes in the second round of screening. The first was to use NMR structures when available, instead of carrying out long MD simulations to extract dominant conformations of the target. The second lesson was to employ more accurate methods in ranking of the compounds, instead of depending solely on the docking scoring function, no matter which combination of scoring methods was used (i.e. average or weighted average binding energies). Hence, in executing the second screening simulation we used 10 NMR target structures, a larger library of compounds (CN chemical library ~100,000 compounds), and the MM-PBSA scoring method to rank the top hits. Eleven compounds from the top hits of this screening were tested experimentally. Five compounds showed positive activity leading to a significantly better-hit rate than the initial round of screening (data not shown). The top hit of the second round of ERCC1 screening is shown in Fig. 6.

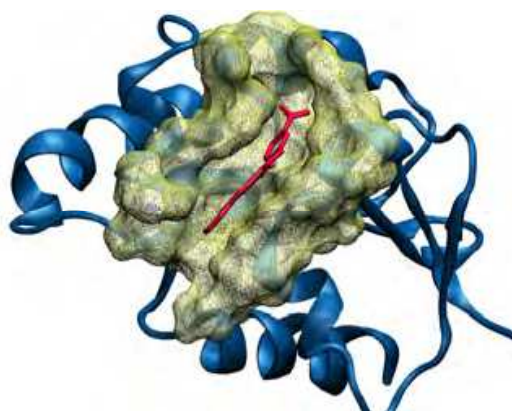


Fig. 6. The top hit from the second screening against the ERCC1 target.

### 3.2 Case II: dual inhibitors of p53-MDM2/4

Over the last two decades, the tumor suppressor protein p53 has been called the “guardian of the genome”. P53 earned this label due to its vital roles in cell cycle, apoptosis, DNA

repair and senescence (Teodoro, et al., 2007). In these processes, p53 responds to cellular stresses, such as hypoxia and DNA damage, by accumulating in the nucleus and activating various pathways to maintain the cell's functional normality (Vogelstein, et al., 2000). As such, tumor cells have developed numerous ways to disable its function. Certainly, the gene *TP53*, which encodes for p53, is mutated or deleted in ~50% of human cancers (Feki & Irminger-Finger, 2004). In the remaining human cancers, while p53 retains its wild type structure, its activity is eradicated by its main cellular inhibitors, murine double minute 2/4 (MDM2/4) proteins (Kubbutat, et al., 1997; Kussie, et al., 1996). Originally, MDM2 was discovered as the main regulator of p53 activity. It acts as an E3 ubiquitin ligase that exports p53 out of the nucleus and promotes its degradation. Moreover, by binding to the transactivation domain of p53 within the nucleus, MDM2 inhibits p53 function as a transcription factor for other proteins. Consequently, over-expression of MDM2 reduces the cellular ability to activate the p53 pathway under stress conditions (Fakharzadeh, et al., 1991).

Structurally related to MDM2, MDM4 (also known as MDMX or HDMX) is a second cellular regulator of p53 (Shvarts, et al., 1996). Although MDM4 lacks the intrinsic E3 ligase activity of MDM2, current models suggest that it acts as a major p53 transcriptional antagonist independent of MDM2 (Toledo, et al., 2006). The binding domains of p53 within MDM2 and MDM4 are very similar (V. Bottger, et al., 1999), offering promise for the discovery of new small molecule compounds that can simultaneously target the two proteins.

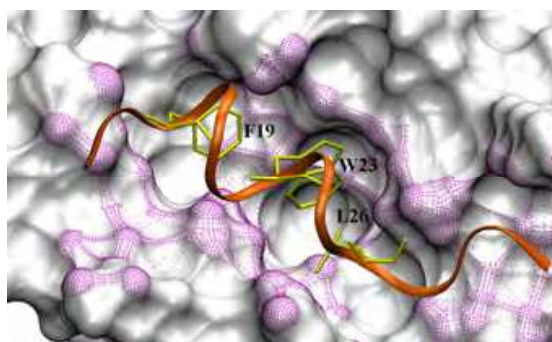


Fig. 7. Illustration of the P53-MDM2 interaction. The p53-binding site within MDM2 (purple) is shown in molecular surface representation with the residues constituting the binding site are highlighted in purple. P53 (orange) is shown in ribbon representation.

The high-resolution crystal structure of the p53-MDM2 complex demonstrated the essential interacting regions located in the MDM2-p53 interface (see Fig. 7) (Kussie, et al., 1996). Essentially, p53 forms an amphipathic-helix peptide (residues 15-29) that is partly buried inside a small but deep, hydrophobic groove on the surface of the MDM2 N-terminal domain (residues 19-102). This interaction involves four key residues from p53, namely F19, L22, W23 and L26 and at least 13 residues from MDM2 (L54, L57, I61, M62, Y67, Q72, V75, F86, F91, V93, I99, Y100 and I103) (A. Bottger, et al., 1997). Interestingly, 10 out of the 13 most important MDM2 residues described above are conserved in MDMX, which indicates that the binding site of p53 within the surface of MDMX is similar to, but not identical with, that of MDM2.

The last decade witnessed the identification of many small-molecule p53-MDM2 inhibitors with promising binding affinities (Patel & Player, 2008). These are analogs of *cis*-imidazoline (Nutlins) (Vassilev, 2004), spiro-oxindole (MI-63 and MI219) (Dastidar, et al., 2008; Ding, et

al., 2006), benzodiazepinedione (TDP665759) (Grasberger, et al., 2005), terphenyl,(L. Chen, et al., 2005), quilinol,(Lu, et al., 2006), chalcone (Stoll, et al., 2001) and sulfonamide (Galatin & Abraham, 2004). Only three compounds, namely, Nutlin-3, MI-219 and TDP665759 showed sufficiently high binding affinity, and desirable pharmacokinetic profiles in cells (Shangary & Wang, 2008) to be seriously considered for clinical development. However, these compounds are more highly selective for MDM2 than for its homolog MDM4. Even nutlin-3 has been shown to be inactive in cancer cells that over-express MDM4 (B. Hu, et al., 2006), opening a new avenue in p53 research and requiring a new generation of MDM2-inhibitors that can target its homolog, MDM4, as well.

3.2.1 Screening against two targets

In our efforts to discover novel compounds that can restore the p53 activity we screened for dual inhibitors of the p53 interactions with both MDM2 and MDM4 (K. Barakat, et al., 2010). Our strategy followed from the evident similarity between the p53 binding sites within the two proteins. We first filtered a subset of our VCC for MDM2 inhibitors Top hits from this search were then screened against the MDM4 target. Compounds that can simultaneously bind to the two targets were considered as potential dual inhibitors. Our VCC-subset included the NCI diversity set, DrugBank compounds. We also enriched the docked compounds with more than 3,168 derivative structures extracted from the known MDM2-inhibitors. This enrichment was obvious, as similar targets are more than likely to bind similar compounds. The initial screening against MDM2 used 28 dominant protein conformations. These conformations represented the apo- and holo-MDM2’s collective conformational dynamics and were extracted from MD simulations, PCA and clustering analyses.

Compound	MDM2 Ranking (kcal/mol)			MDM4 Ranking (kcal/mol)		
	MM-PBSA	AutoDock	Experimental	MM-PBSA	AutoDock	Experimental
MI-219	-10.6 ± 1.5	-9.1 ± 2.2	-11.4	-5.3 ± 1.5	-6.8 ± 2.2	-5.9
Nutlin-3	-9.3 ± 1.3	-8.2 ± 2.2	-9.7	-6.1 ± 1.6	-5.8 ± 2.2	Negligible
TDP665759	-9.5 ± 1.5	-9.1 ± 2.2	-8.4	-5.6 ± 1.4	-8.2 ± 2.2	Negligible
PMI	-10.4 ± 1.6	N/A	-11.6	-12.8 ± 1.5	N/A	-11.5

Table 1. Relative ranking of positive controls using the two scoring methods compared to experimental data.

Scoring of the top MDM2 hits employed two ranking steps. First was a docking-based ranking similar to what was described in the previous ERCC1 study. The objective was to suggest a modest number (300 in this case) of promising hits for the subsequent re-ranking step. This final scoring utilized the MM-PBSA method. Prior to the application of the MM-PBSA method and as described in Fig. 1, all 300 hits were prepared through all-atoms and solvated MD simulations. We also included a recently discovered peptidic MDM2/4 dual inhibitor (PMI) in the rescoring step (Pazgier, et al., 2009). The 300 suggested MDM2 top hits were also docked to the p53-binding site within MDM4 followed by rescoring their binding using MD and MM-PBSA calculations. Table 1 describes the relative ranking of the used positive controls. The apparent IC<sub>50</sub> values for Nutlin3, MI-219, TDP665759 and PMI in binding to MDM2 are 90 nM, 5 nM, 704 nM (Koblish, et al., 2006; Vassilev, 2007) and 3.4 nM35 at 250C, respectively. We did not find explicit values for the binding affinities of the three non-peptide molecules regarding their binding to MDM4, hoverer, it has been



experimentally confirmed that these compounds are weak binders to MDMX (Vassilev, 2007) (Pazgier, et al., 2009).

Although the discrepancy in the MM-PBSA calculations for the interactions of the four inhibitors with MDM2 was about 1 kcal/mol, the predicted values were in an excellent agreement with the experimental data compared to the values obtained by the AutoDock scoring function (see Table 1). These results also illustrated the limitations of AutoDock scoring function in eliminating false positive ligands, i.e. compounds that cannot practically bind but are predicted to bind, from active compounds. For example, the TDP665759 compound was predicted to bind to MDM4 with a relatively high binding energy compared to the rest of the compounds. On the other hand, the MM-PBSA approach selected the real binders for the two protein targets. For MDM2, the four ligands can bind strongly to the protein, while, for MDM4, only the PMI peptide can bind with a very high binding energy.

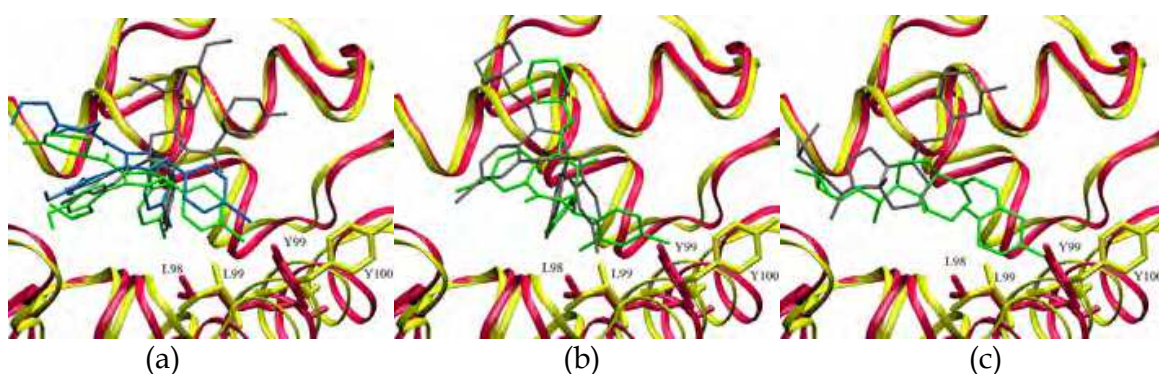


Fig. 8. Structural variations between MDM2 (yellow) and MDMX (red) and their effect on the binding modes of Nutlin-3 (a) and two selected hits from the predicted MDM2/MDM4 inhibitors (b and c). Tyr100 and Leu99 of MDM2 and the same residues in MDM4 are shown in Licorice representations with the same color as that of the two proteins. For each compound, the binding mode within MDM2 is shown in green and within MDM4 is shown in gray. Tyr99 and Leu98 prevent Nutlin-3 from binding to MDM4 with the same binding conformation adopted by Nutlin-2 within the MDM2-pocket (blue). The conformation of nutlin-2 was extracted from the MDM2-nutlin crystal structure 1RV1. On the other hand, compounds Pub#11952782 (b) and ZINC04629876 (c) from the suggested MDM2/MDM4 inhibitor list can tolerate the structural variations in the two binding sites in order to maximize their interactions with the proteins.

Not only did this study reveal a number of promising hits that can simultaneously bind to the two targets, but it also explained why known MDM2-inhibitors such as Nutlin 3 could not bind to MDM4. Although the two binding sites are fairly similar, the MDM4 pocket seemed to be more compact than that of MDM2. This was mainly due to the three residues Pro95, Ser96 and Pro97 in MDM4 that have been replaced by His96, Arg97 and Lys98 in MDM2 (see Fig. 8).

These substitutions are located on one of the alpha helices that comprise the p53 binding site within the two proteins. Consequently, the proline residues (Pro95 and Pro97) in MDM4 shifted this helical domain in MDM4 relative to MDM2 and caused Lys98 and Tyr99 to protrude into the p53-binding cleft within MDM4, making it shallower and less accessible to many of the MDM2 top hits we found. Moreover, we noticed very minor differences in the electrostatic potential distributions around the surfaces of the two proteins (data not



shown), where MDM2 was more positively charged in certain regions deeply located within the binding site. These slight variations in both shape and electrical properties of the two proteins played a considerable role in governing the final conformation adopted by the ligands.

This observation is clear when comparing the binding modes of nutlin within the two pockets (see Fig. 8-a). While Tyr100 and Leu99 of MDM2 extend the binding site allowing nutlin to intimately bind to MDM2, the same residues in MDM4 clash with the drug preventing it from taking the normal conformation that was adopted within MDM2. On the other hand, Fig. 8b-c show how two compounds from the discovered set of proposed MDM2/MDM4 inhibitors were able to tolerate the structural variations between the two binding sites.

### 3.3 Case III: inhibitors of DNA polymerase beta

DNA polymerase beta (pol $\beta$ ), the smallest naturally occurring DNA polymerase enzyme, belongs to the X-family of DNA polymerases (Uchiyama, et al., 2009). DNA pol $\beta$  is a vital member of the base excision repair (BER) pathway (Beard & Wilson, 2006). The enzyme plays a significant role in chemotherapeutic agent resistance, as its over-expression reduces the efficacy of anticancer drug therapies including bleomycin (Parsons, et al., 2004), monofunctional alkylating agents (Liu, et al., 2002), cisplatin (Hoffmann, et al., 1996), and other platinum-based compounds.. Furthermore, small-scale studies on different types of cancer showed that pol $\beta$  is mutated in approximately 30% of tumors, which in turn reduces pol $\beta$  fidelity in DNA synthesis exposing the genome to serious and often deleterious mutations (Chan, et al., 2007; Starcevic, et al., 2004). Based on these findings, pol $\beta$ , the error-prone polymerase of BER, has been seriously considered as a promising therapeutic target for cancer treatment.

Many inhibitors of DNA pol $\beta$  have been identified during the last two decades. To name but a few, this list includes polypeptides (Husain, et al., 1995), fatty acids (Mizushina, et al., 1996), triterpenoids (Tanaka, et al., 1998), sulfolipids (Mizushina, et al., 1998), polar lipids (Ogawa, Murate, Izuta, et al., 1998), secondary bile acids (Ogawa, Murate, Suzuki, et al., 1998), phenalenone-derivatives (Perpelescu, et al., 2002), anacardic acid (J. Z. Chen, Y.; Wang, L.; Sucheck, S.; Snow, A.; Hecht, S., 1998), harbinatic acid (Deng, et al., 1999), flavanoid derivatives (Maloney, et al., 2005), and pamoic acid (H. Y. Hu, et al., 2004). However, most of these inhibitors are either not potent enough or lack sufficient specificity to eventually become approved drugs. Among these compounds pamoic acid (PA) was one of the few compounds that had promising activity against pol $\beta$  and a well-defined binding mode. The compound was initially discovered by Hu and his co-workers (H. Y. Hu, et al., 2004) Their NMR analysis revealed that PA binds to the 8-kDa domain of pol $\beta$  and suggested that the binding pocket is located between the two helices: helix-2 and helix-4 of the 8-kDa domain. Interestingly, the same region has been recognized in different studies to be essential in the DNA binding and deoxyribose phosphate lyase activities of the enzyme (Pelletier, et al., 1994). The precise interactions between PA and the lyase domain of pol $\beta$  were further investigated in a different study (Hazan, et al., 2008) which used a combination protocol of blind docking and NMR analysis to confirm the earlier findings of Hu et al (H. Y. Hu, et al., 2004).

### 3.3.1 Screening against the lyase active site of pol $\beta$

Following the procedure described in Fig. 1, we focused our search space on the binding site of PA, using it as a positive control. Our aim was to discover more potent drug candidates through filtering a library of ~12,500 compounds against 11 protein structures. The molecules tested included the NCI diversity set, the DrugBank set of small-molecules and more than 9,000 fragment structures with drug-like properties extracted from the ZINC database (see section 2.2.1 for a detailed description of these compound libraries). The top 300 hits that showed strong affinity for pol $\beta$  have been validated and rescored using a more robust scoring function, the MM-PBSA method.

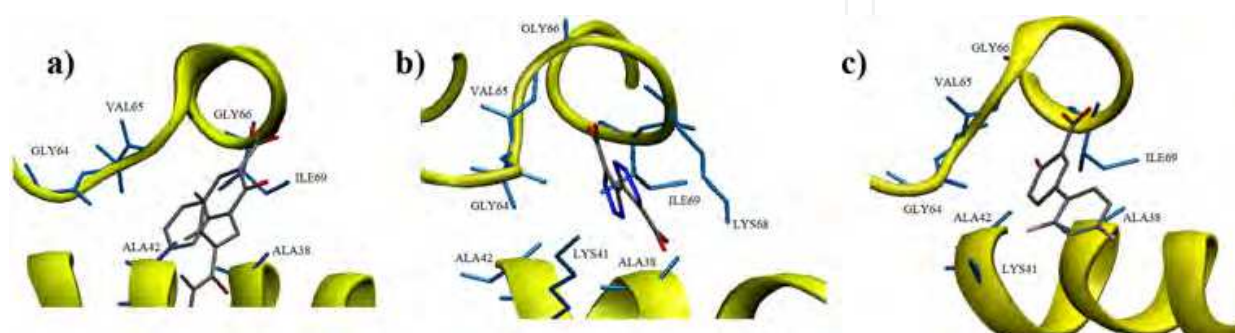


Fig. 9. Binding modes of selected hits. Pol $\beta$  is shown in yellow, important protein residues are shown in blue, and the different atoms of the bound compounds are shown by their representative colors (carbon in gray, oxygen in red, nitrogen in blue and hydrogen in white).

The reported  $K_D$  value for PA binding to pol $\beta$  is 9  $\mu$ M (H. Y. Hu, et al., 2004). Using the AutoDock scoring function, we obtained a value of -6.2 kcal/mol as an estimate for this binding energy. Although this value is in excellent agreement with the experimental measurement (-6.9 kcal/mol) as calculated using the  $K_D$  value, based on the previously described studies, docking scoring functions were not efficient in discriminating false positives in VS experiments and are biased toward their training set of compounds (Tondi, et al., 1999). Consequently, in this work, the top 300 hits were rescored using the MM-PBSA method in order to validate their docking results and confirm their binding to the protein. Fig. 9 demonstrates the binding modes of the top three hits of the MM-PBSA ranking. Similarly to a substantial number of our suggested top hits, the shown compounds are small in size, however, they occupy a considerable portion of the DNA-binding pocket. These lead compounds can be employed as the basis for a further fragment-based drug design step, in order to construct potent and more specific pol  $\beta$  inhibitors.

## 4. Conclusion

DNA repair pathways control the balance between genomic stability's constancy and mutability (Harper & Elledge, 2007). The mode of action of modern anticancer treatments is by inducing damage to the DNA. Over-expression of proteins involved in the DNA repair circuitry boosts the repair activity, removes most of the induced damage and hence, reduces the efficacy of DNA-damaging agents. This unexpected mechanism represents one of the

major factors behind the antitumor drug resistance phenomena observed for these agents (Harper & Elledge, 2007). Therefore, DNA repair proteins are currently considered as valuable targets to improve cancer therapy (Damia & D'Incalci, 2007).

Our group has been involved in *in silico* searches for novel inhibitors of a number of DNA repair proteins. This chapter reviews our efforts in applying computational high throughput screening methods to filter compound libraries for such inhibitors. The chapter contains two main parts. First is a detailed description of all the computational steps that are used in the virtual screening workflow that we follow. Second is a summary of the results we found in applying this protocol to three important DNA repair targets. The three targets are ERCC1-XPA (K. H. Barakat, et al., 2009), an important element of the NER pathway; MDM2 and MDM4 (K. Barakat, et al., 2010), the two cellular inhibitors of p53; and finally DNA pol $\beta$  (K. Barakat, et al., In press); (K. Barakat & Tuszynski, 2011), the error prone polymerase of BER. The aim of this review is to share with the reader our experience in this field from a computational drug discovery perspective. Furthermore, we have attempted to demonstrate that computational tools can be easily applied to DNA repair proteins and eventually arrive at compounds capable of regulating their activities.

## 5. Acknowledgment

Funding for this work was obtained from the Alberta Cancer Foundation, Canadian Breast Cancer Foundation, Alberta's Advanced Education and Technology and the Allard Foundation and NSERC.

## 6. References

- Abagyan, R., & Totrov, M. (2001). High-throughput docking for lead generation. *Curr Opin Chem Biol*, 5(4), 375-382.
- Amadei, A., Linssen, A. B., & Berendsen, H. J. (1993). Essential dynamics of proteins. *Proteins*, 17(4), 412-425.
- Amaro, R. E., Baron, R., & McCammon, J. A. (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J Comput Aided Mol Des*, 22(9), 693-705.
- Barakat, K., Gajewski, M., & Tuszynski, J. (In press). DNA polymerase beta (pol  $\beta$ ) inhibitors: a comprehensive overview.
- Barakat, K., Mane, J., Friesen, D., & Tuszynski, J. (2010). Ensemble-based virtual screening reveals dual-inhibitors for the p53-MDM2/MDMX interactions. *J Mol Graph Model*, 28(6), 555-568.
- Barakat, K., & Tuszynski, J. (2011). Relaxed complex scheme suggests novel inhibitors for the lyase activity of DNA polymerase beta. *J Mol Graph Model*, 29(5), 702-716.
- Barakat, K. H., Torin Huzil, J., Luchko, T., Jordheim, L., Dumontet, C., & Tuszynski, J. (2009). Characterization of an inhibitory dynamic pharmacophore for the ERCC1-XPA interaction using a combined molecular dynamics and virtual screening approach. *J Mol Graph Model*, 28(2), 113-130.
- Beard, W. A., & Wilson, S. H. (2006). Structure and mechanism of DNA polymerase Beta. *Chem Rev*, 106(2), 361-382.

- Bennett, M. J., Barakat, K., Huzil, J. T., Tuszynski, J., & Schriemer, D. C. (2010). Discovery and characterization of the laulimalide-microtubule binding mode by mass shift perturbation mapping. *Chem Biol*, 17(7), 725-734.
- Bottger, A., Bottger, V., Garcia-Echeverria, C., Chene, P., Hochkeppel, H. K., Sampson, W., et al. (1997). Molecular characterization of the hdm2-p53 interaction. *J Mol Biol*, 269(5), 744-756.
- Bottger, V., Bottger, A., Garcia-Echeverria, C., Ramos, Y. F., van der Eb, A. J., Jochemsen, A. G., et al. (1999). Comparative study of the p53-mdm2 and p53-MDMX interfaces. *Oncogene*, 18(1), 189-199.
- Boulikas, T., & Vougiouka, M. (2003). Cisplatin and platinum drugs at the molecular level. (Review). *Oncol Rep*, 10(6), 1663-1682.
- Broughton, H. B. (2000). A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening. *J Mol Graph Model*, 18(3), 247-257, 302-244.
- Carlson, H. A., Masukawa, K. M., Rubins, K., Bushman, F. D., Jorgensen, W. L., Lins, R. D., et al. (2000). Developing a dynamic pharmacophore model for HIV-1 integrase. *J Med Chem*, 43(11), 2100-2114.
- Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., et al. (2005). The Amber biomolecular simulation programs. *J Comput Chem*, 26(16), 1668-1688.
- Chan, K., Houlbrook, S., Zhang, Q. M., Harrison, M., Hickson, I. D., & Dianov, G. L. (2007). Overexpression of DNA polymerase beta results in an increased rate of frameshift mutations during base excision repair. *Mutagenesis*, 22(3), 183-188.
- Chen, J. Z., Y.; Wang, L.; Sucheck, S.; Snow, A.; Hecht, S. (1998). Inhibitors of DNA Polymerase b from *Schoepfia Californica*. *J C S Chem Commun*, 2769-2770.
- Chen, L., Yin, H., Farooqi, B., Sebt, S., Hamilton, A. D., & Chen, J. (2005). p53 alpha-Helix mimetics antagonize p53/MDM2 interaction and activate p53. *Mol Cancer Ther*, 4(6), 1019-1025.
- Damia, G., & D'Incalci, M. (2007). Targeting DNA repair as a promising approach in cancer therapy. *Eur J Cancer*, 43(12), 1791-1801.
- Dastidar, S. G., Lane, D. P., & Verma, C. S. (2008). Multiple peptide conformations give rise to similar binding affinities: molecular simulations of p53-MDM2. *J Am Chem Soc*, 130(41), 13514-13515.
- Davies DL, & Bouldin DW. (1979). A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 1, 224.
- Deng, J. Z., Starck, S. R., Hecht, S. M., Ijames, C. F., & Hemling, M. E. (1999). Harbinatic acid, a novel and potent DNA polymerase beta inhibitor from *Hardwickia binata*. *J Nat Prod*, 62(7), 1000-1002.
- Ding, K., Lu, Y., Nikolovska-Coleska, Z., Wang, G., Qiu, S., Shangary, S., et al. (2006). Structure-based design of spiro-oxindoles as potent, specific small-molecule inhibitors of the MDM2-p53 interaction. *J Med Chem*, 49(12), 3432-3435.
- Fakharzadeh, S. S., Trusko, S. P., & George, D. L. (1991). Tumorigenic potential associated with enhanced expression of a gene that is amplified in a mouse tumor cell line. *EMBO J*, 10(6), 1565-1569.



- Feki, A., & Irminger-Finger, I. (2004). Mutational spectrum of p53 mutations in primary breast and ovarian tumors. *Crit Rev Oncol Hematol*, 52(2), 103-116.
- Galatin, P. S., & Abraham, D. J. (2004). A nonpeptidic sulfonamide inhibits the p53-mdm2 interaction and activates p53-dependent transcription in mdm2-overexpressing cells. *J Med Chem*, 47(17), 4163-4165.
- Garcia, A. E. (1992). Large-amplitude nonlinear motions in proteins. *Phys Rev Lett*, 68(17), 2696-2699.
- Garrett MM, David SG, Robert SH, Ruth H, William EH, Richard KB, et al. (1999). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem.*, 19, 1639.
- Goodsell, D. S., & Olson, A. J. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8(3), 195-202.
- Grasberger, B. L., Lu, T., Schubert, C., Parks, D. J., Carver, T. E., Koblisch, H. K., et al. (2005). Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *J Med Chem*, 48(4), 909-912.
- Harper, J. W., & Elledge, S. J. (2007). The DNA damage response: ten years after. *Mol Cell*, 28(5), 739-745.
- Hazan, C., Boudsocq, F., Gervais, V., Saurel, O., Ciais, M., Cazaux, C., et al. (2008). Structural insights on the pamoic acid and the 8 kDa domain of DNA polymerase beta complex: towards the design of higher-affinity inhibitors. *BMC Struct Biol*, 8, 22.
- Hess, B. (2002). Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys*, 65(3 Pt 1), 031910.
- Hoffmann, J. S., Pillaire, M. J., Garcia-Estefania, D., Lapalu, S., & Villani, G. (1996). In vitro bypass replication of the cisplatin-d(GpG) lesion by calf thymus DNA polymerase beta and human immunodeficiency virus type I reverse transcriptase is highly mutagenic. *J Biol Chem*, 271(26), 15386-15392.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3), 712-725.
- Hu, B., Gilkes, D. M., Farooqi, B., Sebti, S. M., & Chen, J. (2006). MDMX overexpression prevents p53 activation by the MDM2 inhibitor Nutlin. *J Biol Chem*, 281(44), 33030-33035.
- Hu, H. Y., Horton, J. K., Gryk, M. R., Prasad, R., Naron, J. M., Sun, D. A., et al. (2004). Identification of small molecule synthetic inhibitors of DNA polymerase beta by NMR chemical shift mapping. *J Biol Chem*, 279(38), 39736-39744.
- Husain, I., Morton, B. S., Beard, W. A., Singhal, R. K., Prasad, R., Wilson, S. H., et al. (1995). Specific inhibition of DNA polymerase beta by its 14 kDa domain: role of single- and double-stranded DNA binding and 5'-phosphate recognition. *Nucleic Acids Res*, 23(9), 1597-1603.
- Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model*, 45(1), 177-182.
- Kang, S., Ju, W., Kim, J. W., Park, N. H., Song, Y. S., Kim, S. C., et al. (2006). Association between excision repair cross-complementation group 1 polymorphism and clinical



- outcome of platinum-based chemotherapy in patients with epithelial ovarian cancer. *Exp Mol Med*, 38(3), 320-324.
- Koblish, H. K., Zhao, S., Franks, C. F., Donatelli, R. R., Tominovich, R. M., LaFrance, L. V., et al. (2006). Benzodiazepinedione inhibitors of the Hdm2:p53 complex suppress human tumor cell proliferation in vitro and sensitize tumors to doxorubicin in vivo. *Mol Cancer Ther*, 5(1), 160-169.
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc Chem Res*, 33(12), 889-897.
- Kubbutat, M. H., Jones, S. N., & Vousden, K. H. (1997). Regulation of p53 stability by Mdm2. *Nature*, 387(6630), 299-303.
- Kussie, P. H., Gorina, S., Marechal, V., Elenbaas, B., Moreau, J., Levine, A. J., et al. (1996). Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science*, 274(5289), 948-953.
- Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61(4), 704-721.
- Li, L., Elledge, S. J., Peterson, C. A., Bales, E. S., & Legerski, R. J. (1994). Specific association between the human DNA repair proteins XPA and ERCC1. *Proc Natl Acad Sci U S A*, 91(11), 5012-5016.
- Limin, A. L., Dongqing, W., Yixue, L., & Huimin, L. (2011). *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications* (ISBN (ebook)): IGI Publishing.
- Lin, J. H., Perryman, A. L., Schames, J. R., & McCammon, J. A. (2002). Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J Am Chem Soc*, 124(20), 5632-5633.
- Liu, L., Nakatsuru, Y., & Gerson, S. L. (2002). Base excision repair as a therapeutic target in colon cancer. *Clin Cancer Res*, 8(9), 2985-2991.
- Lord, R. V., Brabender, J., Gandara, D., Alberola, V., Camps, C., Domine, M., et al. (2002). Low ERCC1 expression correlates with prolonged survival after cisplatin plus gemcitabine chemotherapy in non-small cell lung cancer. *Clin Cancer Res*, 8(7), 2286-2291.
- Lu, Y., Nikolovska-Coleska, Z., Fang, X., Gao, W., Shangary, S., Qiu, S., et al. (2006). Discovery of a nanomolar inhibitor of the human murine double minute 2 (MDM2)-p53 interaction through an integrated, virtual database screening strategy. *J Med Chem*, 49(13), 3759-3762.
- Maloney, D. J., Deng, J. Z., Starck, S. R., Gao, Z., & Hecht, S. M. (2005). (+)-Myristinin A, a naturally occurring DNA polymerase beta inhibitor and potent DNA-damaging agent. *J Am Chem Soc*, 127(12), 4140-4141.
- McGuire, W. P., & Ozols, R. F. (1998). Chemotherapy of advanced ovarian cancer. *Semin Oncol*, 25(3), 340-348.
- Mizushina, Y., Tanaka, N., Yagi, H., Kurosawa, T., Onoue, M., Seto, H., et al. (1996). Fatty acids selectively inhibit eukaryotic DNA polymerase activities in vitro. *Biochim Biophys Acta*, 1308(3), 256-262.

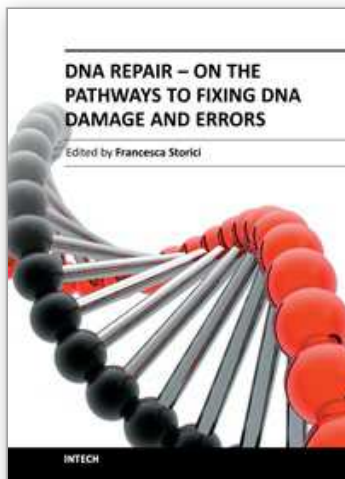
- Mizushina, Y., Watanabe, I., Ohta, K., Takemura, M., Sahara, H., Takahashi, N., et al. (1998). Studies on inhibitors of mammalian DNA polymerase alpha and beta: sulfolipids from a pteridophyte, *Athyrium niponicum*. *Biochem Pharmacol*, 55(4), 537-541.
- Ogawa, A., Murate, T., Izuta, S., Takemura, M., Furuta, K., Kobayashi, J., et al. (1998). Sulfated glycolipid from archaebacterium inhibits eukaryotic DNA polymerase alpha, beta and retroviral reverse transcriptase and affects methyl methanesulfonate cytotoxicity. *Int J Cancer*, 76(4), 512-518.
- Ogawa, A., Murate, T., Suzuki, M., Nimura, Y., & Yoshida, S. (1998). Lithocholic acid, a putative tumor promoter, inhibits mammalian DNA polymerase beta. *Jpn J Cancer Res*, 89(11), 1154-1159.
- Parsons, J. L., Dianova, II, & Dianov, G. L. (2004). APE1 is the major 3'-phosphoglycolate activity in human cell extracts. *Nucleic Acids Res*, 32(12), 3531-3536.
- Patel, S., & Player, M. R. (2008). Small-molecule inhibitors of the p53-MDM2 interaction for the treatment of cancer. *Expert Opin Investig Drugs*, 17(12), 1865-1882.
- Pazgier, M., Liu, M., Zou, G., Yuan, W., Li, C., Li, J., et al. (2009). Structural basis for high-affinity peptide inhibition of p53 interactions with MDM2 and MDMX. *Proc Natl Acad Sci U S A*, 106(12), 4665-4670.
- Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science*, 264(5167), 1891-1903.
- Perpelescu, M., Kobayashi, J., Furuta, M., Ito, Y., Izuta, S., Takemura, M., et al. (2002). Novel phenalenone derivatives from a marine-derived fungus exhibit distinct inhibition spectra against eukaryotic DNA polymerases. *Biochemistry*, 41(24), 7610-7616.
- Rabik, C. A., & Dolan, M. E. (2007). Molecular mechanisms of resistance and toxicity associated with platinating agents. *Cancer Treat Rev*, 33(1), 9-23.
- Rosenberg, E., Taher, M. M., Kuemmerle, N. B., Farnsworth, J., & Valerie, K. (2001). A truncated human xeroderma pigmentosum complementation group A protein expressed from an adenovirus sensitizes human tumor cells to ultraviolet light and cisplatin. *Cancer Res*, 61(2), 764-770.
- Schneider, G., & Bohm, H. J. (2002). Virtual screening and fast automated docking methods. *Drug Discov Today*, 7(1), 64-70.
- Shangary, S., & Wang, S. (2008). Targeting the MDM2-p53 interaction for cancer therapy. *Clin Cancer Res*, 14(17), 5318-5324.
- Shao J, Tanner SW, Thompson N, & Cheatham TE. (2007). Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation*. 3, 2312.
- Shao, J., Tanner, S., Thompson, N., & Cheatham, T. (2007). Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms. *Journal of Chemical Theory and Computation*(3), 2312.
- Shirota, Y., Stoecklacher, J., Brabender, J., Xiong, Y. P., Uetake, H., Danenberg, K. D., et al. (2001). ERCC1 and thymidylate synthase mRNA levels predict survival for colorectal cancer patients receiving combination oxaliplatin and fluorouracil chemotherapy. *J Clin Oncol*, 19(23), 4298-4304.

- Shoichet, B. K., Stroud, R. M., Santi, D. V., Kuntz, I. D., & Perry, K. M. (1993). Structure-based discovery of inhibitors of thymidylate synthase. *Science*, 259(5100), 1445-1450.
- Shvarts, A., Steegenga, W. T., Riteco, N., van Laar, T., Dekker, P., Bazuine, M., et al. (1996). MDMX: a novel p53-binding protein with some functional properties of MDM2. *EMBO J*, 15(19), 5349-5357.
- Stahura, F. L., & Bajorath, J. (2004). Virtual screening methods that complement HTS. *Comb Chem High Throughput Screen*, 7(4), 259-269.
- Starcevic, D., Dalal, S., & Sweasy, J. B. (2004). Is there a link between DNA polymerase beta and cancer? *Cell Cycle*, 3(8), 998-1001.
- Stoll, R., Renner, C., Hansen, S., Palme, S., Klein, C., Belling, A., et al. (2001). Chalcone derivatives antagonize interactions between the human oncoprotein MDM2 and p53. *Biochemistry*, 40(2), 336-344.
- Tanaka, N., Kitamura, A., Mizushima, Y., Sugawara, F., & Sakaguchi, K. (1998). Fomitelic acids, triterpenoid inhibitors of eukaryotic DNA polymerases from a basidiomycete, fomitella fraxinea. *J Nat Prod*, 61(9), 1180.
- Teodoro, J. G., Evans, S. K., & Green, M. R. (2007). Inhibition of tumor angiogenesis by p53: a new role for the guardian of the genome. *J Mol Med*, 85(11), 1175-1186.
- Toledo, F., Krummel, K. A., Lee, C. J., Liu, C. W., Rodewald, L. W., Tang, M., et al. (2006). A mouse p53 mutant lacking the proline-rich domain rescues Mdm4 deficiency and provides insight into the Mdm2-Mdm4-p53 regulatory network. *Cancer Cell*, 9(4), 273-285.
- Tondi, D., Slomczynska, U., Costi, M. P., Watterson, D. M., Ghelli, S., & Shoichet, B. K. (1999). Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chem Biol*, 6(5), 319-331.
- Tsodikov, O. V., Ivanov, D., Orelli, B., Staeresincic, L., Shoshani, I., Oberman, R., et al. (2007). Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J*, 26(22), 4768-4776.
- Uchiyama, Y., Takeuchi, R., Kodera, H., & Sakaguchi, K. (2009). Distribution and roles of X-family DNA polymerases in eukaryotes. *Biochimie*, 91(2), 165-170.
- Vassilev, L. T. (2004). Small-molecule antagonists of p53-MDM2 binding: research tools and potential therapeutics. *Cell Cycle*, 3(4), 419-421.
- Vassilev, L. T. (2007). MDM2 inhibitors for cancer therapy. *Trends Mol Med*, 13(1), 23-31.
- Vogelstein, B., Lane, D., & Levine, A. J. (2000). Surfing the p53 network. *Nature*, 408(6810), 307-310.
- Warren, G. L., Andrews, C. W., Capelli, A. M., Clarke, B., LaLonde, J., Lambert, M. H., et al. (2006). A critical assessment of docking programs and scoring functions. *J Med Chem*, 49(20), 5912-5931.
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., et al. (2006). DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res*, 34(Database issue), D668-672.

Wood, R. D., Mitchell, M., Sgouros, J., & Lindahl, T. (2001). Human DNA repair genes. *Science*, 291(5507), 1284-1289.

IntechOpen

IntechOpen



## **DNA Repair - On the Pathways to Fixing DNA Damage and Errors**

Edited by Dr. Francesca Storici

ISBN 978-953-307-649-2

Hard cover, 380 pages

**Publisher** InTech

**Published online** 09, September, 2011

**Published in print edition** September, 2011

DNA repair is fundamental to all cell types to maintain genomic stability. A collection of cutting-edge reviews, DNA Repair - On the pathways to fixing DNA damage and errors covers major aspects of the DNA repair processes in a large variety of organisms, emphasizing foremost developments, questions to be solved and new directions in this rapidly evolving area of modern biology. Written by researchers at the vanguard of the DNA repair field, the chapters highlight the importance of the DNA repair mechanisms and their linkage to DNA replication, cell-cycle progression and DNA recombination. Major topics include: base excision repair, nucleotide excision repair, mismatch repair, double-strand break repair, with focus on specific inhibitors and key players of DNA repair such as nucleases, ubiquitin-proteasome enzymes, poly ADP-ribose polymerase and factors relevant for DNA repair in mitochondria and embryonic stem cells. This book is a journey into the cosmos of DNA repair and its frontiers.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Barakat, K. and Tuszyński, J. (2011). Virtual Screening for DNA Repair Inhibitors, DNA Repair - On the Pathways to Fixing DNA Damage and Errors, Dr. Francesca Storici (Ed.), ISBN: 978-953-307-649-2, InTech, Available from: <http://www.intechopen.com/books/dna-repair-on-the-pathways-to-fixing-dna-damage-and-errors/virtual-screening-for-dna-repair-inhibitors>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen